Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells

Cell Reports
Report

Transient Accumulation of 5-Carboxylcytosine Indicates Involvement of Active Demethylation in Lineage Specification of Neural Stem Cells

Lee M. Wheldon,[1,6] Abdulkadir Abakir,[2,6] Zoltan Ferjentsik,[3,6] Tatiana Dudnakova,[4] Stephanie Strohbuecker,[2] Denise Christie,[3] Nan Dai,[5] Shengxi Guan,[5] Jeremy M. Foster,[5] Ivan R. Corrêa, Jr.,[5] Matthew Loose,[3] James E. Dixon,[2] Virginie Sottile,[2,*] Andrew D. Johnson,[3] and Alexey Ruzov[2,*]

[1]Medical Molecular Sciences, Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK
[2]Division of Cancer and Stem Cells, School of Medicine, Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[3]School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK
[4]School of Biological Sciences, University of Edinburgh, King's Buildings, Mayfield Road, Edinburgh EH9 3JR, UK
[5]New England Biolabs, Inc., 240 County Road, Ipswich, MA 01938, USA
[6]Co-first author
*Correspondence: virginie.sottile@nottingham.ac.uk (V.S.), alexey.ruzov@nottingham.ac.uk (A.R.)
http://dx.doi.org/10.1016/j.celrep.2014.05.003

SUMMARY

5-Methylcytosine (5mC) is an epigenetic modification involved in regulation of gene activity during differentiation. Tet dioxygenases oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC). Both 5fC and 5caC can be excised from DNA by thymine-DNA glycosylase (TDG) followed by regeneration of unmodified cytosine via the base excision repair pathway. Despite evidence that this mechanism is operative in embryonic stem cells, the role of TDG-dependent demethylation in differentiation and development is currently unclear. Here, we demonstrate that widespread oxidation of 5hmC to 5caC occurs in postimplantation mouse embryos. We show that 5fC and 5caC are transiently accumulated during lineage specification of neural stem cells (NSCs) in culture and in vivo. Moreover, 5caC is enriched at the cell-type-specific promoters during differentiation of NSCs, and TDG knockdown leads to increased 5fC/5caC levels in differentiating NSCs. Our data suggest that active demethylation contributes to epigenetic reprogramming determining lineage specification in embryonic brain.

INTRODUCTION

5-Methylcytosine (5mC) is a DNA modification contributing to the regulation of gene activity during development and differentiation (Reik et al., 2001; Bird, 2002). Tet (Ten-eleven translocation) dioxygenases (TET1/2/3) can oxidize 5mC, generating 5-hydroxymethylcytosine (5hmC), which, according to a growing body of evidence, plays specific biological roles in embryonic stem cells (ESCs) and the mammalian brain (Tahiliani et al., 2009; Ficz et al., 2011; Lister et al., 2013). 5hmC can be enzymatically oxidized further to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Ito et al., 2011; He et al., 2011). Both 5fC and 5caC can be recognized and excised from DNA by thymine-DNA glycosylase (TDG) followed by subsequent regeneration of unmodified cytosine by the components of base excision repair (BER) pathway (He et al., 2011; Maiti and Drohat, 2011). Correspondingly, recent genome-wide analyses of 5fC/5caC distribution in mouse ESCs (mESCs) revealed a TDG-dependent accumulation of these marks in gene regulatory elements, suggesting that this mechanism of active demethylation is operative in mESCs (Shen et al., 2013; Song et al., 2013).

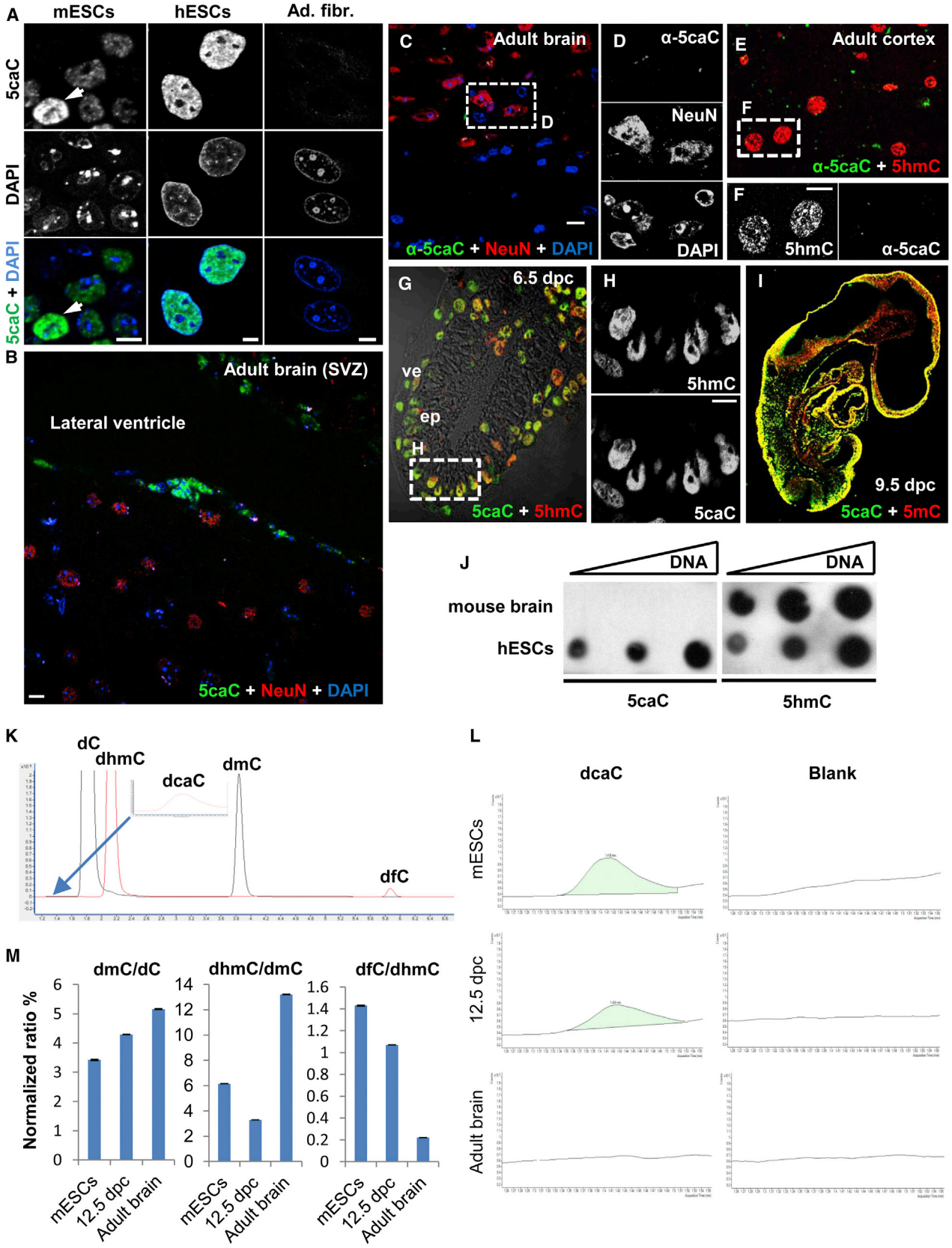
Paradoxically, contrary to the mESC-based results, recent studies imply that a passive replication-dependent demethylation mechanism is involved in epigenetic reprogramming during development of primordial germ cells (PGCs) (Seisenberger et al., 2012; Hackett et al., 2013). Likewise, although 5fC and 5caC are detectable in mouse zygotes, these marks are diluted in a replication-dependent manner during preimplantation development instead of being actively removed from DNA (Inoue et al., 2011). Thus, the role of BER-dependent demethylation in cellular differentiation and development is currently unclear.

In this study, we aimed to examine if a TDG/BER-dependent active DNA demethylation pathway is employed during mammalian embryogenesis and differentiation.

RESULTS

Widespread Oxidation of 5hmC to 5caC Occurs in ESCs and in Mouse Postimplantation Embryos

Since 5fC and 5caC can serve as intermediates in active demethylation, we checked whether these marks are detectable in ESCs and postimplantation embryos employing a sensitive immunostaining method based on the use of peroxidase-conjugated secondary antibody and tyramide amplification reagent, which we have previously applied to 5hmC detection (Ruzov



(legend on next page)

et al., 2011). Using a 5caC antibody (Active Motif), which specifically discriminates this mark from other forms of modified cytosine (Inoue et al., 2011), we performed 5caC immunostaining on mESCs, human ESCs (hESCs), and mouse embryonic sections using human adult-derived fibroblasts and adult mouse brain tissue as controls (Figures 1A–1I). Whereas hESCs and a subpopulation of mESCs were highly positive for 5caC, it was undetectable in the fibroblast cultures (Figure 1A). Likewise, although some cells in the subventricular zone (SVZ) of the brain displayed a nonnegligible 5caC staining (Figure 1B), we could not detect it in the hippocampus or cortex of adult brain (Figures 1C–1F; Figure S1A). In contrast, 5caC was present in both epiblast and visceral endoderm of embryos at 6.5 days postcoitum (dpc) and in a range of embryonic tissues at 9.5 and 12.5 dpc (Figures 1G–1I; Figure S1B). We confirmed the specificity of this staining in competition experiments and in dot blots (Figure 1J; Figures S1C and S1D).

Although we observed elevated levels of 5fC signal in the same contexts as 5caC enrichment, the intensity of 5fC staining was always considerably weaker than that of 5caC (Figures S1E and S1F). Since 5fC antibody was at least ten times less sensitive than 5caC in dot blots with modified DNA standards (Figure S1G), the weak intensity of 5fC staining was likely due to the relatively low sensitivity of the corresponding antibody.

To confirm our results, we performed mass spectrometry (MS) detection of 5caC and 5fC in mESCs, embryonic head tissue at 12.5 dpc, and the adult brain samples. Although 5caC signal was not strong enough for the accurate MS quantification, in agreement with our antibody-based results, we could detect this mark by MS in ESCs and in 12.5-dpc embryonic head samples but not in the DNA isolated from the adult mouse brain (Figures 1K and 1L; Figure S1H). MS quantification of 5fC in the same samples revealed that the 5fC/5hmC ratio was also significantly lower in the adult brain DNA compared with that derived from 12.5 dpc embryonic tissue or ESCs (Figure 1M).

Whereas the intensity of 5caC immunostaining signal was fairly uniform in hESCs, it varied widely between different cells in a mESC colony, ranging from undetectable (54.2% cells) to intensities comparable to that observed in hESCs (12.5% cells) (Figures S2A–S2D). Correspondingly, the levels of 5caC differed several fold between mouse and human ESCs in DNA dot blots (Figure S2E). Although the percentage of 5caC highly positive mESCs appeared higher in 2i medium than in serum-containing medium (Figure S2F), patterns of 5caC distribution were similar for both culturing conditions. Likewise, mESCs synchronized in early S-phase retained the heterogeneity of 5caC signal (Figure S2G). We also did not detect any correlation between the intensity of 5caC staining and the expression levels of Oct4 or Nanog in mESCs (Figure S2H). Notably, in both hESCs and mESCs, 5caC was predominantly euchromatic, displaying a significantly higher degree of colocalization with 5hmC than with 5mC (Figures S3A–S3J).

Dynamics of the Spatial Distribution of 5caC and 5hmC in Mouse Embryonic Brain

Because 5caC was detectable in embryonic brain, but not in most of the adult brain cells (Figures 1B–1F; Figure S1B), we assessed the distribution of both 5hmC and 5caC during several stages of embryonic brain development. In the adult brain, 5hmC was considerably more enriched in NeuN-expressing postmitotic neurons than in glial cells (Figures S4A and S4B). In contrast, most of the brain cells 12.5 dpc exhibited 5hmC staining of fairly equal strength (Figures S4C and S4D). This pattern of 5hmC distribution changed at 13.5 dpc where, similar to adult brain, 5hmC staining intensity was variable between different cellular populations (Figures S4C–S4F). This alteration in the patterns of 5hmC distribution corresponds to a period when active specification of neuronal and glial precursors is underway (Liu et al., 2002; Delaunay et al., 2008) and correlates with a gradual increase in mRNA levels of *Tet2*, de novo methyltransferase *Dnmt3a*, and neuronal markers (*β-III tubulin*) and with a concomitant reduction of stem cell markers (e. g. *Sox2*) and *Dnmt3b* mRNAs (Figures S4G–S4I). Notably, the expression of BER pathway components, *Tdg* and *PARP1*, peaks at the same time point, between 9.5 and 13.5 dpc (Figure S4J).

5caC staining intensity in brain cells was relatively low at 11.5 dpc, peaking at 12.5 dpc, and then dropping by 70% at 13.5 dpc (Figures 2A and 2B). This corresponded to a decrease of 5caC:5hmC colocalization (from 40% to 3%–5%) during developmental progression (Figure 2C). DNA dot blots confirmed the immunostaining results (Figures 2D and S2E). 5caC and 5hmC were distributed in a semi-inverse manner in brain cells 12.5–13.5 dpc, with 5caC staining becoming more restricted by 13.5 dpc (Figures 2E–2I). Notably, whereas the majority of

Figure 1. 5caC Is Detectable in ESCs and Postimplantation Embryos

(A) 5caC signal in mESCs, hESCs, and human adult-derived fibroblasts (Ad. fibr.). Samples were processed using identical experimental conditions and imaged at the same settings. Single channels and merge views are shown.

(B–D) 5caC costaining with NeuN in the SVZ (B) and in a region of adult brain localized at the border of gray and white matter, shown in (C) and (D).

(E and F) 5caC costaining with 5hmC in a representative region of adult brain cortex.

(G and H) 5caC and 5hmC signals in mouse embryo 6.5 dpc. Ep, epiblast; ve, visceral endoderm.

(I) 5caC/5mC immunostaining of a sagittal section of mouse embryo 9.5 dpc.

Merge views are shown in (B), (C), (E), (G), and (I); single channels for the areas marked with dotted squares are shown in (D, (F), and (H). The slides shown in (E), (F), and (G) through (I) were processed in parallel and imaged at the same settings. Scale bars, 10 μm.

(J) DNA dot blot of 5caC and 5hmC in hESCs and adult mouse brain tissue.

(K) Overlaid of total ion chromatogram of global dC, dmC, dhmC, dfC, and dcaC content recorded in dynamic multiple reaction monitoring mode. Red line profile represents approximately 250 times more concentrated DNA than the black line profile. The insert shows a 1,000× expansion of the mass count axis in the dcaC region.

(L) dcaC peaks in mESCs, head tissue 12.5 dpc, and adult brain DNA samples accompanied with blank runs shown in the same scale. The amount of DNA analyzed in mESC samples were about three times higher than in the other samples.

(M) dmC/dC, dhmC/dmC, and dfC/dhmC ratios obtained from the quantification of MS peaks in mESCs, head tissue 12.5 dpc, and adult brain DNA samples. See also Figures S1–S3.

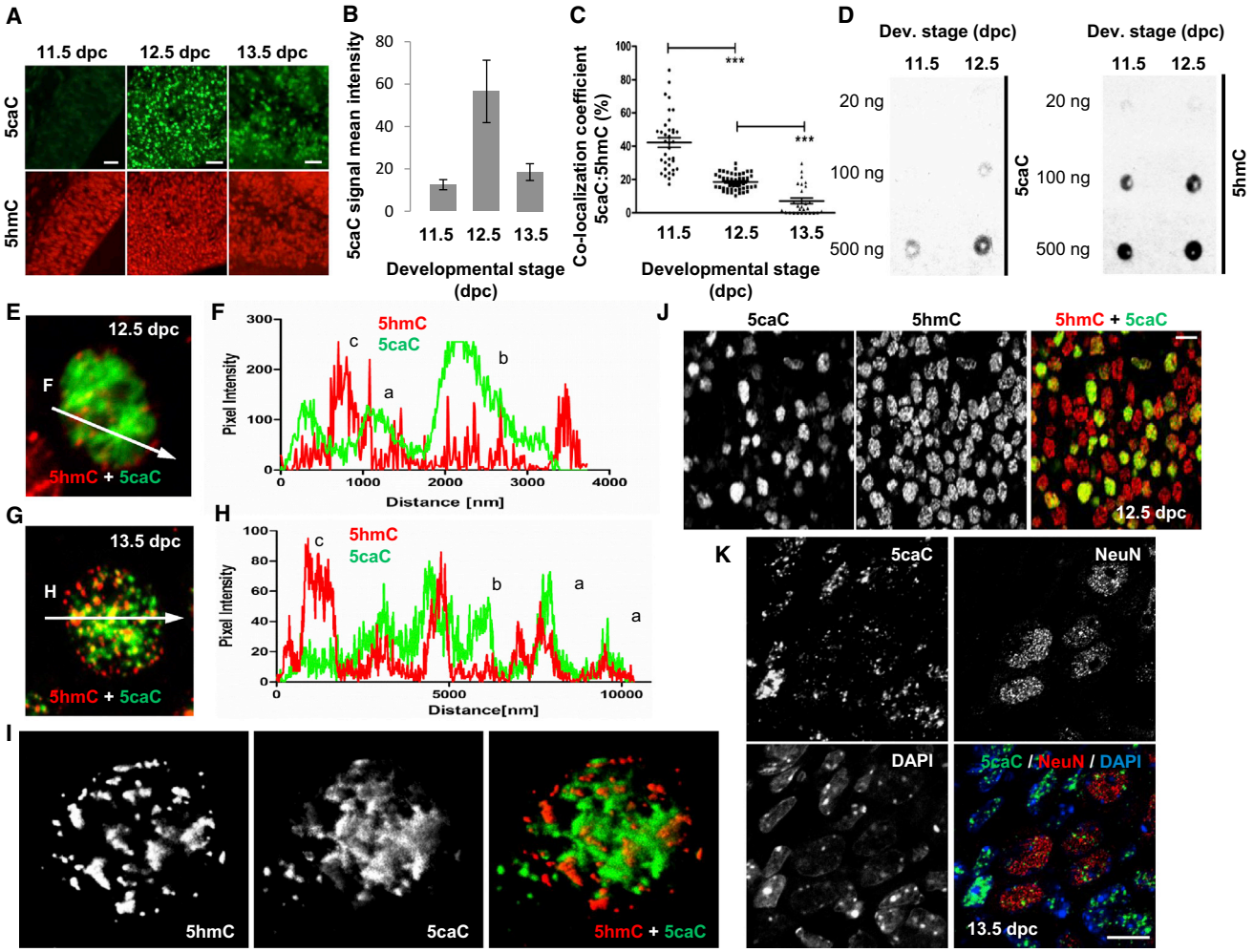


Figure 2. 5caC Is Transiently Accumulated in the Embryonic Brain between 11.5 and 13.5 dpc

(A) 5caC/5hmC immunostaining of equivalent forebrain regions analogous to the ones presented in Figure S1C at 11.5, 12.5, and 13.5 dpc. Sections were processed using identical experimental conditions and imaged at the same settings.

(B) 5caC signal intensity quantification in the sections shown in (A). Data are mean signal intensities ± SEM.

(C) 5caC:5hmC colocalization coefficient values for the equivalent region of embryonic forebrain at 11.5, 12.5, and 13.5 dpc. ***$p < 0.0001$.

(D) DNA dot blot of 5caC and 5hmC in embryonic head tissue 11.5 and 12.5 dpc. The amounts of DNA loaded on to membranes are indicated.

(E–H) Merge views in (E) and (G) and the profiles of 5hmC and 5caC signal intensities in (F) and (H) demonstrate the pattern of 5hmC and 5caC nuclear distribution in brain cells 12.5 and 13.5 dpc. 5caC intensity follows that of 5hmC in some regions (a), whereas in other areas, high 5caC signal corresponds to low 5hmC intensity (b) or strong 5hmC staining parallels weak 5caC signal (c).

(I) 3D reconstruction of 5hmC and 5caC signals distribution in a nucleus of a forebrain cell 13.5 dpc.

(J) 5caC and 5hmC staining in forebrain 12.5 dpc.

(K) 5caC codetection with NeuN in brain 13.5 dpc.

Single channels and merge views are shown in (I) through (K).

Scale bars, 25 μm in (A) and 10 μm in (J) and (K).

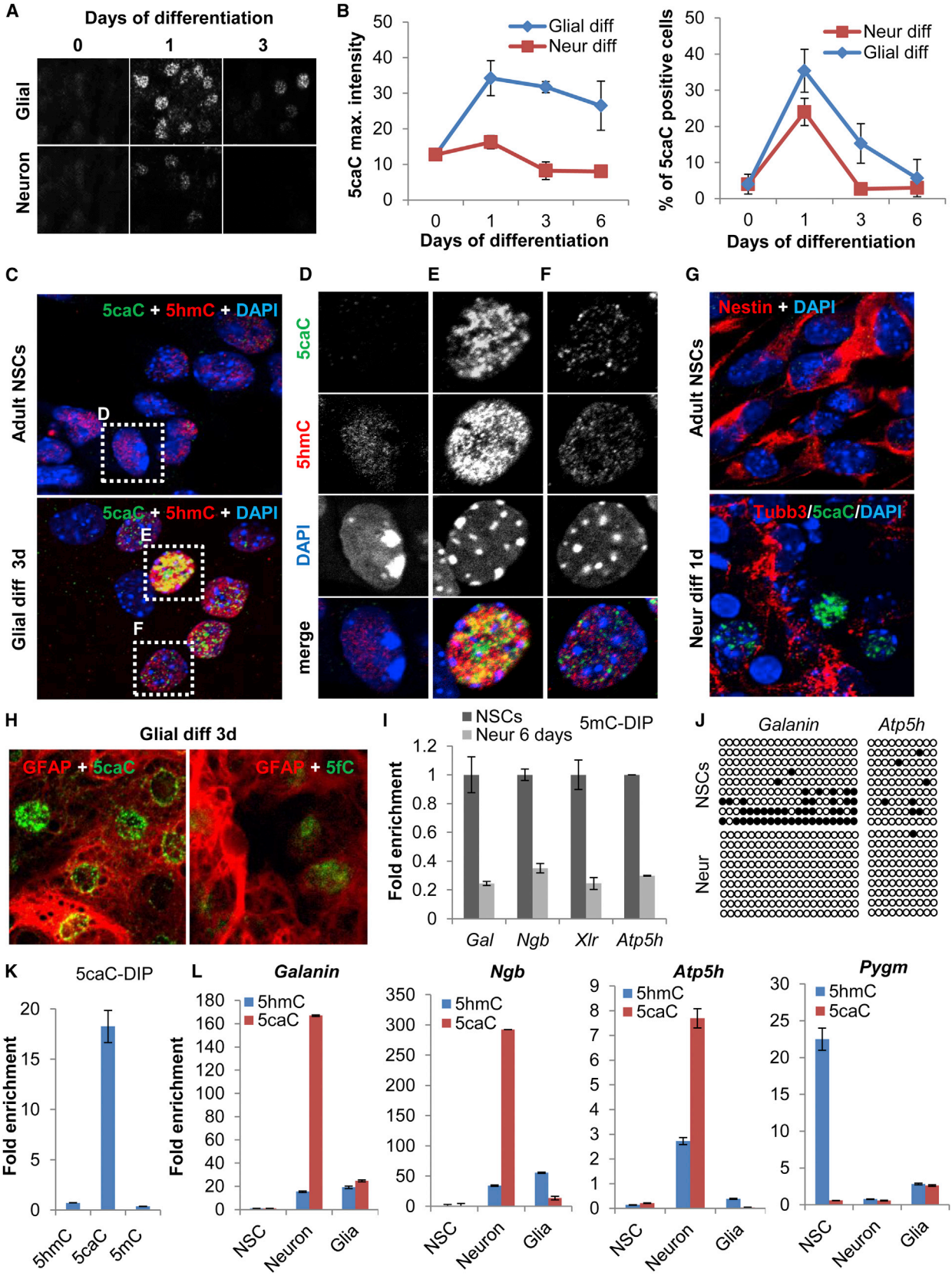
See also Figure S1C.

brain cells 12.5 dpc exhibited 5hmC signal of similar strength, the 5caC staining intensity varied between them from strong to virtually undetectable (Figure 2J). Moreover, levels of 5caC signal were lower in NeuN-expressing cells at 13.5 dpc, indicating that the degree of 5hmC oxidation in embryonic brain may be cell type specific (Figure 2K).

Thus, the cell populations of embryonic brain start to differ in their 5hmC content between 12.5 and 13.5 dpc, which parallels with temporary accumulation of 5caC at certain regions of their genome.

5fC and 5caC Are Transiently Accumulated during Neuronal and Glial Differentiation of Neural Stem Cells

To identify the 5caC-enriched cell types, we examined the levels of this mark during differentiation of adult neural stem cells (NSCs) to neuronal and glial lineages in vitro. Although 5caC



(legend on next page)

was undetectable in the majority of undifferentiated NSCs, strongly 5caC-positive cells started to appear in NSC cultures at initial stages of glial and neuronal differentiation simultaneously, with cells expressing early neuronal (β-III tubulin) and glial (glial fibrillary acidic protein; GFAP) markers (Figures 3A–3H). This 5caC accumulation was transient, with highest numbers of 5caC-positive cells at days 1–3 after induction of differentiation (Figures 3A and 3B). The 5caC dynamics and staining intensity differed between the two differentiation regimes, with markedly higher levels of 5caC signal observed during glial differentiation (Figures 3A and 3B). Similar to embryonic brain (Figures 2D–2H), 5caC and 5hmC exhibited semi-inverse patterns of nuclear distribution in differentiating NSCs (Figures 3A and 3C–3F). Although 5fC was also detectable at the same stages of glial differentiation, its staining intensity was considerably weaker than that of 5caC (Figure 3H).

A number of promoters of genes involved in adult neurogenesis were reported to be hypermethylated in *Tet1* knockout (KO) mice (Zhang et al., 2013). Our 5mC DNA immunoprecipitation (DIP) and bisulfite sequencing experiments revealed that several of these promoters exhibited decreases in 5mC levels during neural differentiation of NSCs (Figures 3I and 3J). To test if this demethylation involved oxidation of 5hmC to 5caC, we used 5hmC/5caC DIP. In agreement with a previous study (Shen et al., 2013), the 5caC antibody specifically recognized this mark in DIP with synthetically modified DNA fragments (Figure 3K). While 5caC was effectively undetectable at tested promoters in undifferentiated NSCs, it was enriched at the promoters of *Galanin*, *Neuroglobin* (*Ngb*), and *Atp5h* in the cells differentiating into neuronal, but not into glial, lineages (Figure 3L).

Collectively, these results indicate that 5caC transiently accumulates at cell-type-specific genomic regions during the initial stages of the differentiation of NSCs.

TDG Knockdown Leads to an Increase in 5fC/5caC during Glial Differentiation

Since the 5caC accumulation was transient, to determine whether TDG is involved in elimination of this mark from DNA, we performed RNA-interference-mediated TDG knockdown (siTDG) in fetal NSCs, followed by their differentiation and subsequent analysis of 5caC distribution (Figure 4A). siTDG treatment resulted in approximately 75% downregulation of TDG expression compared to NSCs transfected with nontargeting small interfering RNA (siControl) (Figure 4B). Neither siTDG nor siControl undifferentiated NSCs exhibited any significant 5caC (Figures 4C and 4D) or 5fC (data not shown) staining. In line with previous observations that *Tdg* KO neuronal progenitor cells failed to differentiate to neurons due to rapid loss of cell viability (Cortázar et al., 2011), siTDG resulted in massive cell death during the initial 2 days of neuronal differentiation, complicating any analysis of 5caC distribution. However, although most siControl cells exhibited weak or absent 5fC/5caC staining at day 3 of glial differentiation, we observed a significant increase in the number of highly 5fC/5caC-positive cells following siTDG, with a high proportion of 5caC-positive GFAP-expressing cells, suggesting that TDG is involved in elimination of this mark from DNA during differentiation (Figures 4E and 4F).

Given that, using DIP, we detected the accumulation of 5caC at CpG-rich promoter regions of a number of key glial markers (*Olig1*, *Olig2*, *Fgfr1*, *Fgfr4*, *Pdgfr*, and *GFAP*) during glial differentiation (Figure 4G), which corresponded to a demethylation of specific CpGs in differentiated cells (Figure 4H), we examined expression of these genes in siTDG glial cells. Remarkably, some of the corresponding mRNAs were significantly (four to five times for *GFAP* and *Pdgfr)* upregulated on siTDG treatment (Figure 4I), which may indicate a functional link between 5hmC oxidation and transcription akin to that previously reported for mESCs (Raiber et al., 2012).

DISCUSSION

The profiles of 5mC and 5hmC distribution are dynamic during brain development and vary between different brain cell types (Lister et al., 2013; Kriaucionis and Heintz, 2009). Moreover, in fetal brain, 5hmC is enriched at putative regulatory elements that are hypomethylated in the adult brain, suggesting Tet-dependent demethylation of these sequences during development (Lister et al., 2013). In line with this, TET1 has been reported to promote BER-dependent active DNA demethylation in the adult brain via conversion of 5mC to 5hmC (Guo et al., 2011). Furthermore, *Tet1* KO mice are characterized by altered neurogenesis, poor learning, and impaired memory extinction, associated with hypermethylation of promoters of neurogenesis-related genes (Zhang et al., 2013; Rudenko et al., 2013). The demethylation of specific promoters has also been

Figure 3. 5caC Is Transiently Accumulated during Differentiation of Adult NSCs

(A) 5caC staining in adult NSCs differentiating toward glial and neuronal lineages at indicated days postinduction. Samples were processed in parallel and imaged at the same settings.

(B) Results of quantification of 5caC signal maximal intensity and percentages of 5caC-positive cells during initial days of adult NSC differentiation. Data are means ± SEM.

(C–F) Codetection of 5caC with 5hmC in undifferentiated adult NSCs (upper panel) and at day 3 of their glial differentiation (lower panel). Merged views are shown in (C) and single channels for the nuclei marked with dotted rectangles in (D) through (F).

(G) Immunostaining of undifferentiated adult NSCs for nestin and NSCs at day 1 of neuronal differentiation for 5caC and β-III tubulin (Tubb3).

(H) Coimmunostaining of 5caC (left panel) and 5fC (right panel) with GFAP in NSCs at day 3 of glial differentiation. Identical exposure settings were used.

(I) 5mC DIP of indicated promoters in undifferentiated NSCs and in NSCs at day 6 of neuronal differentiation.

(J) Bisulfite sequencing of indicated promoter regions in undifferentiated NSCs and NSCs at day 6 of neuronal differentiation (Neur diff). Filled circles represent protected (5mC or 5hmC), and empty circles represent unprotected (unmodified C) CpGs.

(K) 5caC DIP of 5hmC-, 5caC-, or 5mC-containing oligonucleotides spiked with carrier DNA.

(L) 5hmC and 5caC DIP on indicated promoters in undifferentiated NSCs and in NSCs differentiating toward neuronal (Neuron) and glial (Glia) lineages. Experimental error is expressed as ±SEM.

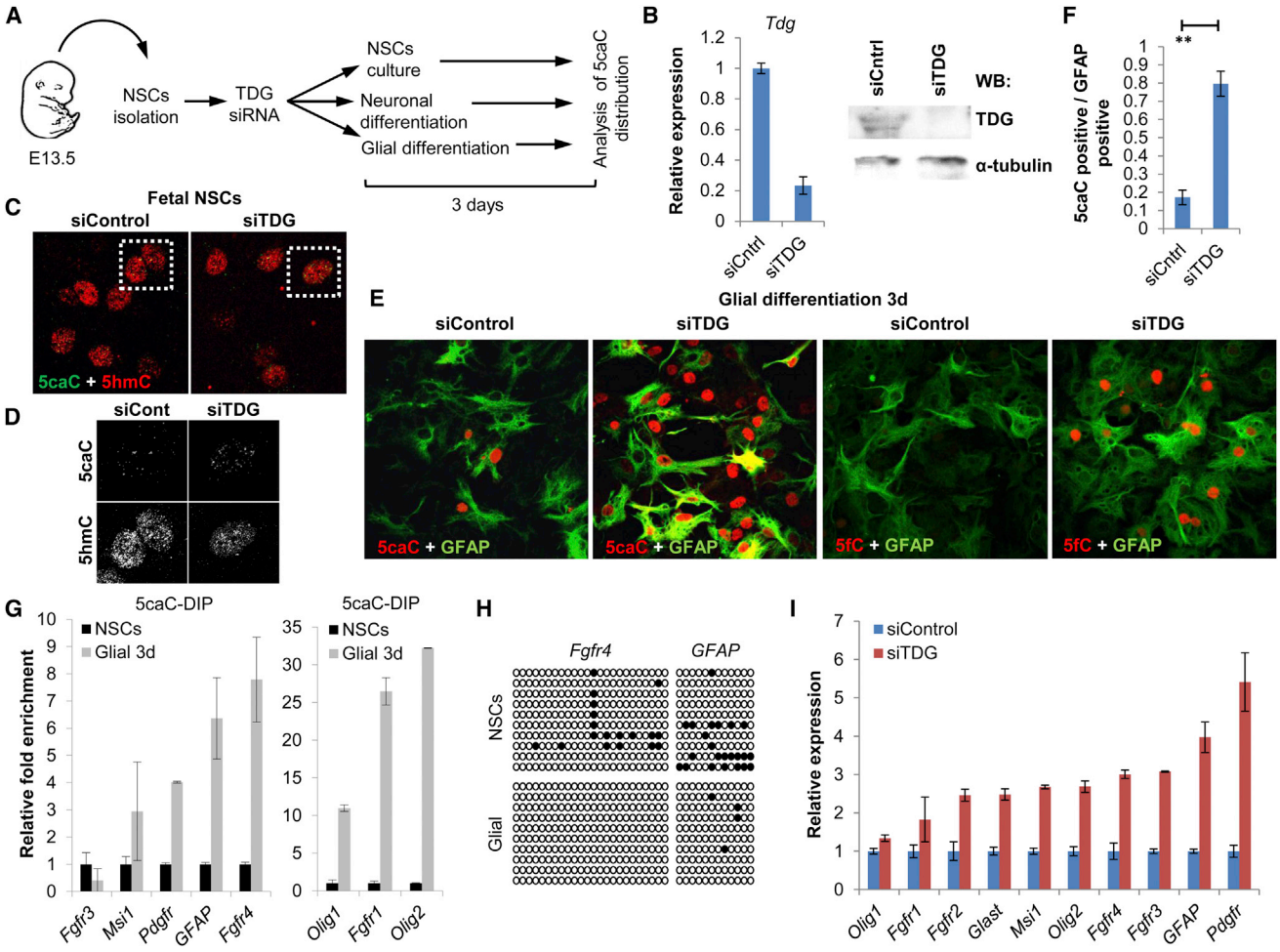


Figure 4. TDG Depletion Leads to Increase in 5fC/5caC Content during Glial Differentiation

(A) Schematic illustrating the design of experiment on TDG depletion in differentiating NSCs.

(B) Expression of TDG in siControl and siTDG NSCs.

(C and D) 5caC/5hmC immunostaining of siControl and siTDG undifferentiated NSCs. Merged views are shown in (C), and single channels for insets are shown in (D).

(E) Coimmunostaining of 5caC (left panels) and 5fC (right panels) with GFAP in siControl and siTDG NSCs at day 3 of glial differentiation. Merged views are shown.

(F) Numbers of 5caC-positive cells normalized to the numbers of GFAP-expressing cells in siControl and siTDG NSCs at day 3 of glial differentiation. Data are means ± SEM. **$p < 0.001$.

(G) Bisulfite sequencing of indicated CpG-rich promoter regions in NSCs and glial cells. Filled circles represent protected (5mC or 5hmC) CpGs; empty circles represent unprotected (unmodified C) CpGs.

(H) 5caC DIP on indicated promoters in undifferentiated NSCs and in NSCs at day 3 of glial differentiation.

(I) Relative expression levels of indicated mRNAs in siTDG and siControl NSCs at day 3 of glial differentiation.

implicated in the mechanisms of neuronal to glial switch as a determinant of astrocyte (Fan et al., 2005; Namihira et al., 2009) and neuronal differentiation (Martinowich et al., 2003). Correspondingly, in the present study, we show that the initial stages of glial and neuronal differentiation are characterized by global accumulation of 5caC, suggesting an involvement of a 5caC/TDG/BER-dependent demethylation in lineage specification of NSCs. In agreement with our results, *Tdg* KO leads to aberrant patterns of DNA methylation and embryonic lethality at 11.5 dpc (Cortellino et al., 2011; Cortázar et al., 2011), the stage immediately preceding the peak of 5caC enrichment in embryonic brain that we observed. Likewise, demethylation of *GFAP* promoter, where we detected 5caC accumulation during glial differentiation, has been reported as a requirement for GFAP expression and astrocyte differentiation (Namihira et al., 2009).

Consistent with the *Tdg* KO study in which epigenetic abnormalities were evident only upon cell lineage commitment (Cortázar et al., 2011), we witnessed an increase in 5caC-positive cells during differentiation but not in undifferentiated siTDG NSCs. This may indicate that 5caC/TDG/BER-dependent demethylation is utilized as a general mechanism for rearrangement of the 5hmC/5mC profiles during terminal differentiation of postmitotic somatic cell types in mammals. In contrast, analogous to

PGCs and preimplantation embryos (Hackett et al., 2013; Inoue et al., 2011), a potentially less mutagenic replication-dependent demethylation mechanism may be operative in actively dividing embryonic cellular populations.

In this respect, the elevated levels of 5caC we observed in a subpopulation of mESCs may indicate the existence of a dynamic stem cell state associated with epigenetic priming of differentiation. Contrasting with mESCs, we observed strong uniform 5caC staining in hESCs. Given that unlike mESC media the KnockOut Serum Replacement used for our hESCs culture contains ascorbic acid (Chung et al., 2010), promoting Tet-mediated 5caC generation (Yin et al., 2013), the high levels of this mark in hESCs may be due to the ascorbic-acid-dependent activation of Tet enzymatic activity occurring in these cells.

EXPERIMENTAL PROCEDURES

ESCs Culture

All the animal-involved procedures were performed in accordance with the University of Nottingham review board. HM1 mESCs were cultured either in serum-containing media or under 2i medium conditions. HUES-7 hESCs were maintained without feeders in conditioned medium containing KnockOut Serum Replacement (Life Technologies).

Derivation, Culture, and Differentiation of NSCs and RNA Interference Knockdown of TDG

The derivation, culture, and differentiation of NSCs were performed as described elsewhere (Alcock and Sottile, 2009). For TDG depletion, fetal NSCs were transfected with 50 pmol of Dharmacon siGENOME siRNA duplexes (Thermo Fisher Scientific) against mouse TDG (Thermo Fisher Scientific, catalog number M-040666-01) and nontargeting siRNA #2 (Thermo Fisher Scientific, catalog number D-001210-02) using Lipofectamine 2000 in antibiotic-free medium. The transfections were repeated after 48 hr. Differentiation was induced 24 hr after the second transfection. Control NSCs were kept in NSC medium. Cells were collected for analysis 72 hr after induction of differentiation.

Immunocytochemistry, Immunohistochemistry, and Dot Blots

These procedures were performed as described elsewhere (Ruzov et al., 2011). Antibodies are listed in the Supplemental Information.

Mass Spectrometry

This was carried out according to the procedures of Hashimoto et al. (2014).

Confocal Microscopy

Images (400–600 nm optical sections) were acquired with a Zeiss LSM 700 AxioObserver confocal microscope using a Plan-Apochromat 63×/1.40 Oil DIC M27 objective and processed using ImageJ, Adobe Photoshop, and ZEN Zeiss LSM 700 imaging software.

5mC, 5hmC, and 5caC DIP

These were carried out according to the procedures of Ficz et al. (2011) and Shen et al. (2013) using mouse monoclonal 5mC antibody (clone 33D3, Diagenode), rabbit polyclonal 5hmC and 5caC antibodies (Active Motif), and magnetic Dynabeads (Invitrogen).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.05.003.

AUTHOR CONTRIBUTIONS

L.M.W. and A.A. performed immunostaining, microscopy, and image analysis; Z.F. contributed to DIP, siRNA analysis, and real-time PCR; N.D., S.G., J.M.F., and I.R.C. performed MS; T.D., D.C., J.E.D., S.S., and V.S. contributed cell lines and reagents; A.R., L.M.W., A.A., Z.F., M.L., V.S., and A.D.J. analyzed the data; A.R. conceived the project and wrote the paper.

ACKNOWLEDGMENTS

We thank Poulam Patel, Ami Ketley, Evangelos Delivopoulos, Rebecca Trueman, Xuewei Qu, Sarwar Abedeen, and Amal Surrati for materials and technical help. This work was supported by the Royal Society (RG110530 to A.R.).

Received: October 9, 2013
Revised: April 4, 2014
Accepted: May 2, 2014
Published: May 29, 2014

REFERENCES

Alcock, J., and Sottile, V. (2009). Dynamic distribution and stem cell characteristics of Sox1-expressing cells in the cerebellar cortex. Cell Res. *19*, 1324–1333.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Chung, T.L., Brena, R.M., Kolle, G., Grimmond, S.M., Berman, B.P., Laird, P.W., Pera, M.F., and Wolvetang, E.J. (2010). Vitamin C promotes widespread yet specific DNA demethylation of the epigenome in human embryonic stem cells. Stem Cells *28*, 1848–1855.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Delaunay, D., Heydon, K., Cumano, A., Schwab, M.H., Thomas, J.L., Suter, U., Nave, K.A., Zalc, B., and Spassky, N. (2008). Early neuronal and glial fate restriction of embryonic neural stem cells. J. Neurosci. *28*, 2551–2562.

Fan, G., Martinowich, K., Chin, M.H., He, F., Fouse, S.D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., et al. (2005). DNA methylation controls the timing of astrogliogenesis through regulation of JAK-STAT signaling. Development *132*, 3345–3356.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398–402.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423–434.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452.

Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Corrêa, I.R., Jr., Zheng, Y., and Cheng, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature *506*, 391–395.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303–1307.

Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. Cell Res. *21*, 1670–1676.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300–1303.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929–930.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Liu, Y., Wu, Y., Lee, J.C., Xue, H., Pevny, L.H., Kaprielian, Z., and Rao, M.S. (2002). Oligodendrocyte and astrocyte development in rodents: an in situ and immunohistological analysis during embryonic development. Glia *40*, 25–43.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chem. *286*, 35334–35338.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890–893.

Namihira, M., Kohyama, J., Semi, K., Sanosaka, T., Deneen, B., Taga, T., and Nakashima, K. (2009). Committed neuronal precursors confer astrocytic potential on residual neural precursor cells. Dev. Cell *16*, 245–255.

Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome Biol. *13*, R69.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089–1093.

Rudenko, A., Dawlaty, M.M., Seo, J., Cheng, A.W., Meng, J., Le, T., Faull, K.F., Jaenisch, R., and Tsai, L.H. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron *79*, 1109–1122.

Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G., et al. (2011). Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res. *21*, 1332–1342.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell *48*, 849–862.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692–706.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678–691.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930–935.

Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. J. Am. Chem. Soc. *135*, 10396–10403.

Zhang, R.R., Cui, Q.Y., Murai, K., Lim, Y.C., Smith, Z.D., Jin, S., Ye, P., Rosa, L., Lee, Y.K., Wu, H.P., et al. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell *13*, 237–245.

Cell Reports, Volume *7*

Supplemental Information

Transient Accumulation of 5-Carboxylcytosine Indicates Involvement of Active Demethylation in Lineage Specification of Neural Stem Cells

Lee M. Wheldon, Abdulkadir Abakir, Zoltan Ferjentsik, Tatiana Dudnakova, Stephanie Strohbuecker, Denise Christie, Nan Dai, Shengxi Guan, Jeremy M. Foster, Ivan R. Corrêa, Jr., Matthew Loose, James E. Dixon, Virginie Sottile, Andrew D. Johnson, and Alexey Ruzov

Figure S1

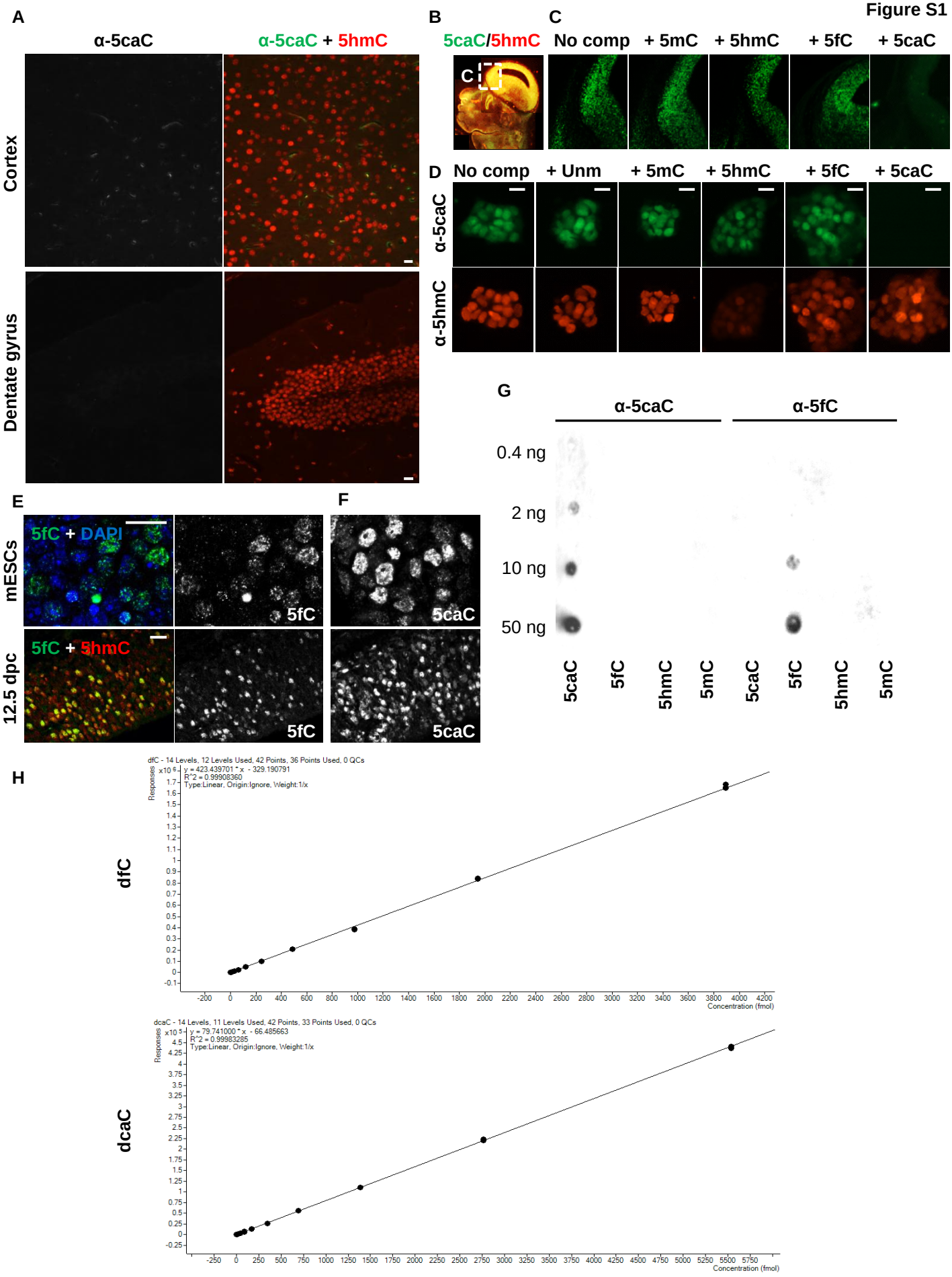

A
α-5caC
α-5caC + 5hmC
Cortex
Dentate gyrus
B
5caC/5hmC
C
C
No comp
+ 5mC
+ 5hmC
+ 5fC
+ 5caC
D
No comp
+ Unm
+ 5mC
+ 5hmC
+ 5fC
+ 5caC
α-5caC
α-5hmC
E
F
mESCs
12.5 dpc
5fC + DAPI
5fC
5caC
5fC + 5hmC
5fC
5caC
G
α-5caC
α-5fC
0.4 ng
2 ng
10 ng
50 ng
5caC
5fC
5hmC
5mC
5caC
5fC
5hmC
5mC
H
dfC
dfC - 14 Levels, 12 Levels Used, 42 Points, 36 Points Used, 0 QCs
y = 423.439701 * x - 329.190791
R^2 = 0.99908360
Type:Linear, Origin:Ignore, Weight:1/x
Responses
Concentration (fmol)
dcaC
dcaC - 14 Levels, 11 Levels Used, 42 Points, 33 Points Used, 0 QCs
y = 79.741000 * x - 66.485663
R^2 = 0.99983285
Type:Linear, Origin:Ignore, Weight:1/x
Responses
Concentration (fmol)

Figure S1 related to Figures 1 and 2. Anti-5caC antibody specifically recognizes 5caC in mouse pre-implantation embryos and ESCs. (A) The results of immunostaining of adult mouse brain sections using anti-5hmC and anti-5caC antibodies. Representative views of the cortex and dentate gyrus part of the hippocampus are shown. Individual channels for anti-5caC antibody and merge views are presented. (B-D) 5caC immunostaining disappears in the presence of 5caC- but not 5mC-, 5hmC-, 5fC-enriched or unmodified DNA. (B) 5hmC and 5caC co-immunostaining of a sagittal section of mouse 12.5 dpc embryo head. Merge view is shown. The position of a region equivalent to the ones shown in (C) is marked with dashed rectangle. (C) Immunostaining of serial lateral sections of mouse 12.5 dpc embryonic forebrain region using anti-5caC antibody in the presence of 5mC-, 5hmC-, 5fC- or 5caC-enriched DNA (+5mC DNA, +5hmC DNA, +5fC DNA and +5caC DNA correspondingly) and without any DNA competitor (no comp.). Single channels for 5caC are presented. (D) Immunostaining of mouse ESCs using anti-5caC and anti-5hmC antibodies in the presence of unmodified, 5mC-, 5hmC-, 5fC- or 5caC-enriched DNA (Unmod., +5-mC DNA, +5-hmC DNA, +5-fC DNA and +5-caC DNA correspondingly) and without any DNA competitor (No comp.). Single channels for 5caC and for 5hmC are presented. (E) Co-detection of 5fC with DAPI in mESCs (upper panel) and immunostaining of 5fC and 5hmC in a lateral section of mouse embryonic forebrain at 12.5 dpc (lower panel). Merge views and single channels for 5fC are shown. (F) 5caC signal in mESCs and in 12.5 dpc embryonic forebrain sections processed in parallel (on the same chamber slides/slides) using identical times of incubation with tyramide to the samples shown in (E). Single channels for 5caC are presented. The 5fC in (E) and 5caC in (F) were visualized using identical confocal settings. Scale bars are 20 μm. (G) The results of dot-blots with equal dilutions (0.4-50 ng) of 5caC- , 5fC-, 5hmC- and 5mC-containing DNA fragments hybridized using anti-5caC or anti-5fC antibodies. (H) External calibration curves for d5caC and d5fC. The concentration of the stock solution for

each standard nucleoside was determined by UV, and diluted to desired concentration range (0.5 – 4000 fmol). Calibration curves were obtained by measuring each standard nucleoside in triplicate at 14 different concentrations.

Figure S2

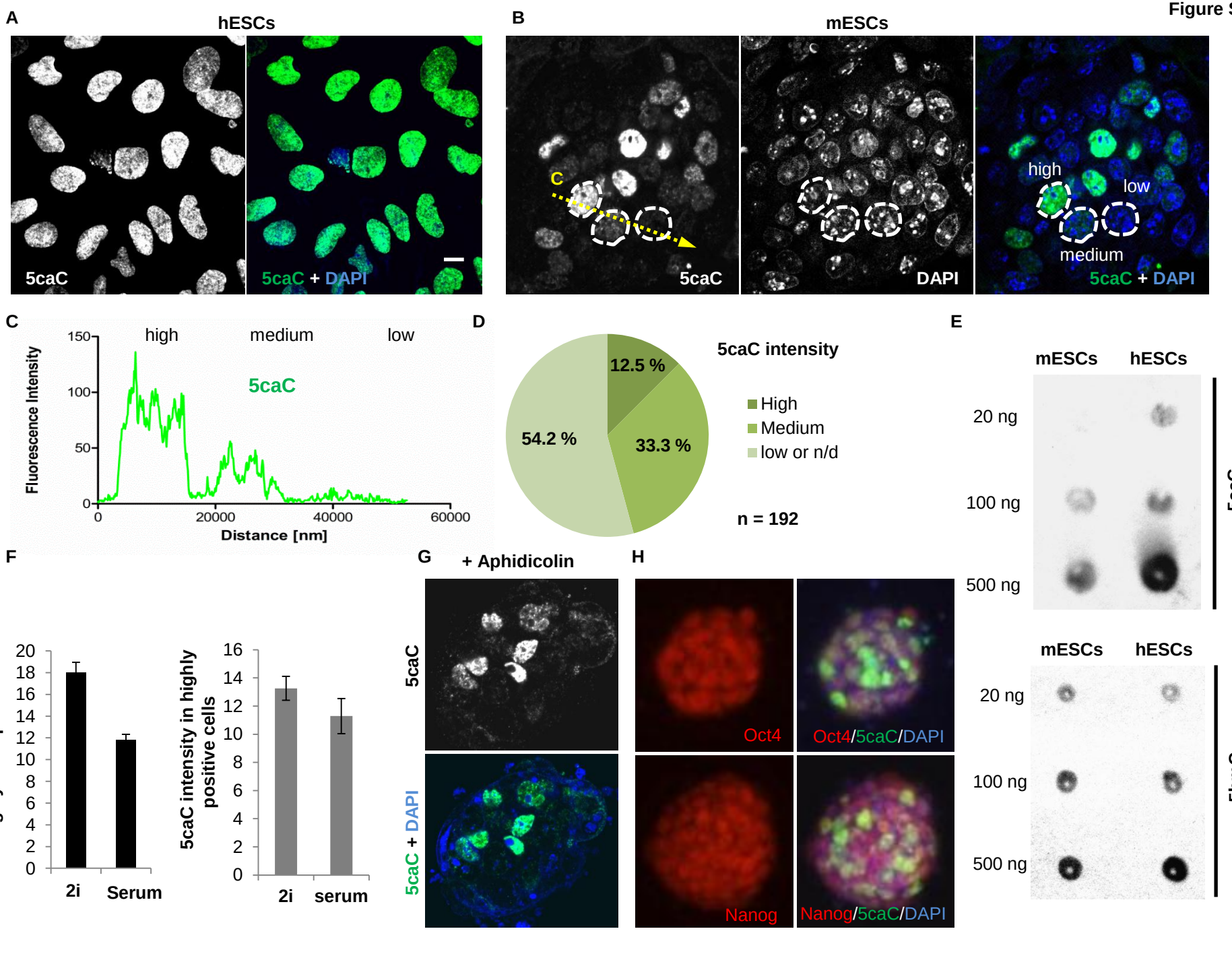

A
hESCs
5caC
5caC + DAPI
B
mESCs
C
5caC
DAPI
high
low
medium
5caC + DAPI
C
high
medium
low
5caC
Fluorescence Intensity
150
100
50
0
0
20000
40000
60000
Distance [nm]
D
5caC intensity
12.5 %
33.3 %
54.2 %
High
Medium
low or n/d
n = 192
E
mESCs
hESCs
20 ng
100 ng
500 ng
5caC
mESCs
hESCs
20 ng
100 ng
500 ng
5hmC
F
% of highly 5caC positive cells
20
18
16
14
12
10
8
6
4
2
0
2i
Serum
5caC intensity in highly positive cells
16
14
12
10
8
6
4
2
0
2i
serum
G
+ Aphidicolin
5caC
5caC + DAPI
H
Oct4
Oct4/5caC/DAPI
Nanog
Nanog/5caC/DAPI

Figure S2 related to Figure 1. mESCs are heterogeneous in their 5caC levels. (A, B) Co-detection of 5caC with DAPI in a monolayer culture of HUES7 hESCs (A) and in a colony of mESCs (B). Examples of cells with different levels of 5caC signal (designated as "high", "medium" or "low" – with maximum FI: 75-200, 25-75 and <25, respectively) used for the categorization of 5caC staining presented in (D) are shown. The position of the region used for generating the signal intensity profile shown in (C) is marked with a yellow dotted arrow. 5caC channels (as well as DAPI channel for mESCs) and merged views are presented. (C) 5caC signal intensity profile across the region of three mESCs nuclei with high, medium or low levels of this modification demonstrated the marked variation (50- to 100-fold) between different cells in the mESCs colony. (D) Pie-chart showing the proportion of cells with different levels of 5caC staining based on quantification of 5 representative images of mESC colonies. (E) The results of dot blot assays with differing amounts (20, 100 and 500 ng) of genomic DNA isolated from mouse and human ESCs hybridized using anti-5caC and anti-5hmC antibodies. (F) The percentage of 5caC-highly positive cells in mESC colonies cultured in serum-containing media and under 2i conditions. 40 mESC colonies were analysed for each condition. The results of 5caC-signal quantification from individual 5caC-highly positive cells in serum- and 2i-cultured mESCs are shown in parallel. Mean values of mean intensities of the 5caC signal in 40 individual highly 5caC-positive cells are shown for each condition. (G) Co-detection of 5caC with DAPI in a representative colony of mESCs synchronized at G1/S boundary with aphidicolin. 5caC channel and merged view are shown. (H) Co-immunostaining of 5caC with Oct4 and Nanog in mESCs colonies cultured in the presence of serum. Single channels for Oct4/Nanog and merged views are shown.

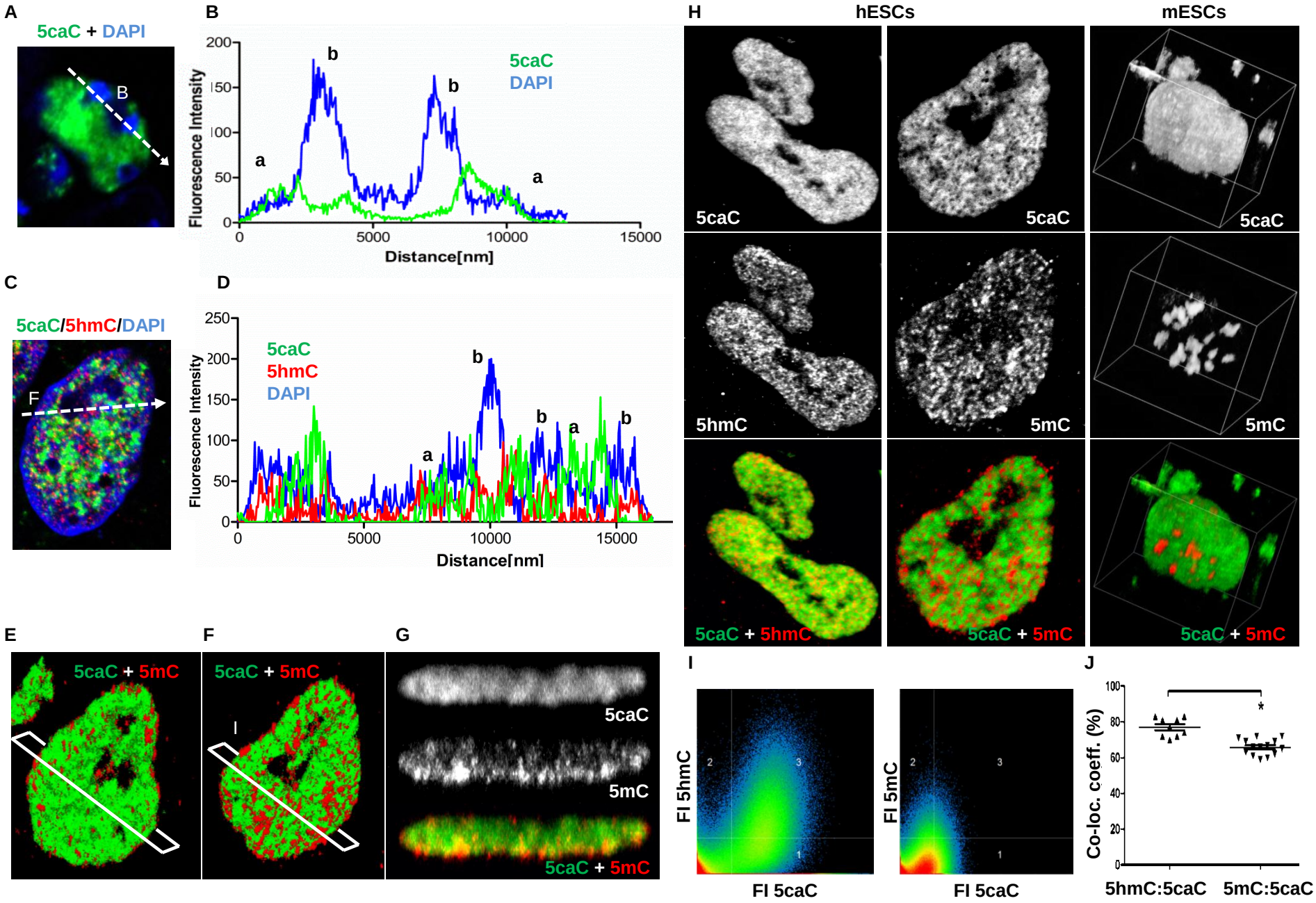
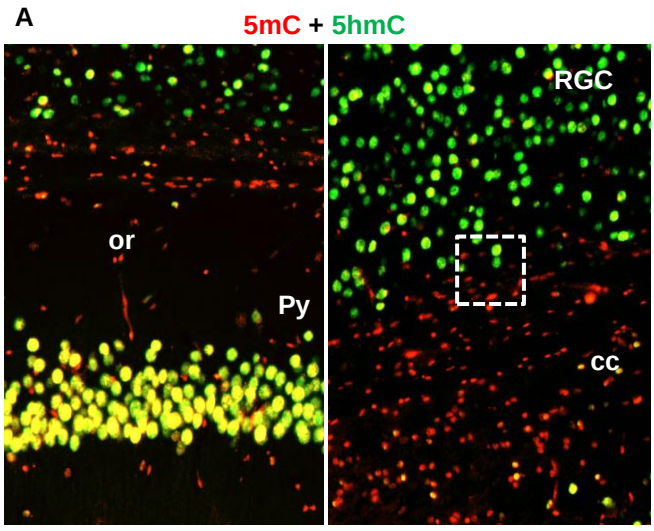
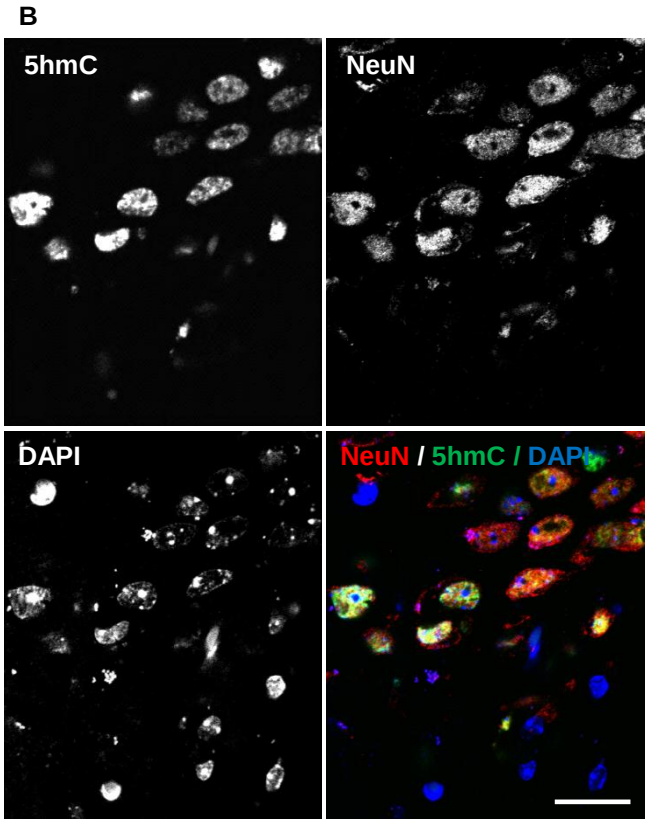
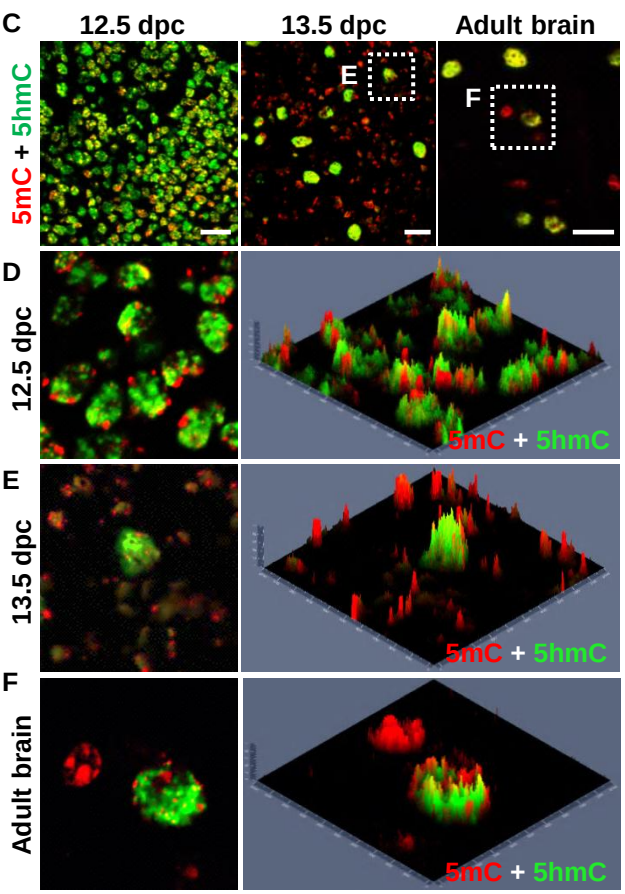

A
5caC + DAPI
B
B
5caC
DAPI
Fluorescence Intensity
Distance[nm]
a
b
C
5caC/5hmC/DAPI
F
D
5caC
5hmC
DAPI
Fluorescence Intensity
Distance[nm]
E
5caC + 5mC
F
5caC + 5mC
I
G
5caC
5mC
5caC + 5mC
H
hESCs
mESCs
5caC
5hmC
5mC
5caC + 5hmC
5caC + 5mC
I
FI 5hmC
FI 5caC
FI 5mC
FI 5caC
J
Co-loc. coeff. (%)
5hmC:5caC
5mC:5caC

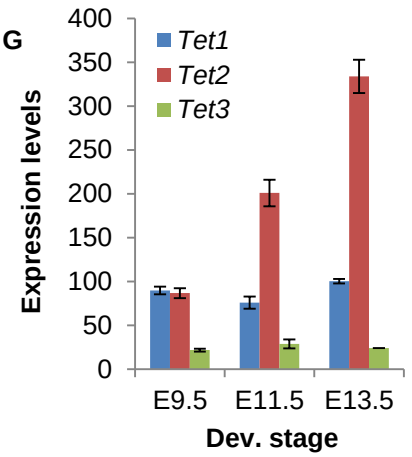
Figure S3 related to Figure 1. 5caC is localized to euchromatin in ESCs nuclei exhibiting high degree of co-localization with 5hmC but not with 5mC. In highly 5caC-positive mESCs (A, B) 5caC signal intensity generally follows DAPI intensity in the nuclear regions where DAPI staining is low or moderate ("a") and drops in heterochromatic nuclear areas with strong DAPI signal ("b"). Single channels and merge views are shown. The region of the nuclus used for generation of corresponding signal intensity profile shown in (B) is marked with dotted arrow. (C) Similar to 5caC, 5hmC is localized to euchromatic nuclear regions in hESCs. Single channels and merge views are shown. 5caC, 5hmC and DAPI signal intensities profile across the region of the nucleus marked with a dotted line are presented in (D). (E, F) 3D reconstruction of 5caC and 5mC distribution at the top and bottom surfaces of a single representative hESCs nucleus is shown. The intensity of 5caC signal is relatively uniform throughout hESCs nuclei, whereas 5mC exhibits a punctate pattern of nuclear distribution. 5mC speckles are predominantly localized in regions adjacent to the nuclear periphery and their concentration is obviously higher near the bottom surfaces of the nuclei (F) compared to their top surfaces (E). The position of the transverse section of the nucleus shown in (G) is marked with a white rectangle. Single channels and merge views are shown in (G). (H, I) 5caC and 5hmC display similar spatial distribution patterns in hESC nuclei with signal distribution plots indicating a markedly higher degree of co-localization between these two marks than between 5caC and 5mC. (H) 3D reconstructions of 5caC/5hmC and 5caC/5mC signal distribution within representative hESC nuclei (left and central panels) compared with 3D reconstruction of 5caC/ 5mC signals distribution inside a single mESCs nucleus with high level of 5caC staining. Single channels and merge views are shown. (I) 5caC/ 5mC and 5caC/5hmC FI (fluorescence intensity) co-localisation plots for representative optical sections of hESCs nuclei depicted in (H). Lines represent threshold intensities (FI = 50). (J)

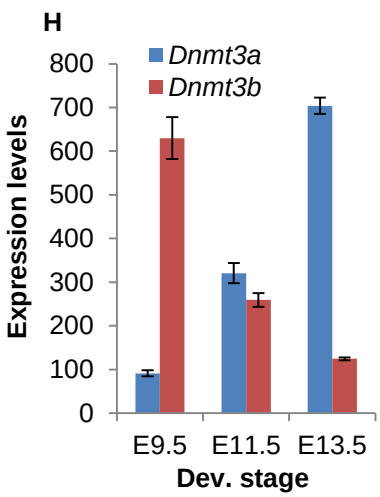
5hmC:5caC co-localization coefficient is significantly higher than that of 5mC:5caC in hESCs, *$p<0.05$.

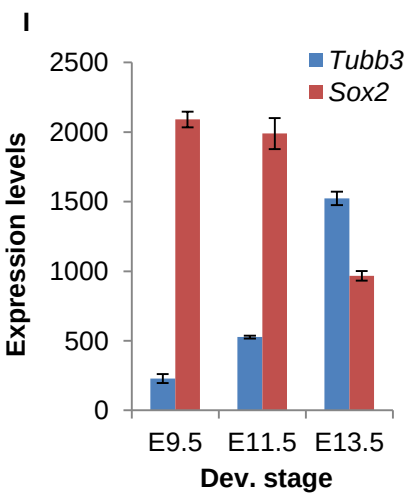

A
5mC + 5hmC
or
Py
RGC
cc


B
5hmC
NeuN
DAPI
NeuN / 5hmC / DAPI


C
12.5 dpc
13.5 dpc
Adult brain
5mC + 5hmC
E
F
D
12.5 dpc
5mC + 5hmC
E
13.5 dpc
5mC + 5hmC
F
Adult brain
5mC + 5hmC


G
Tet1
Tet2
Tet3
Expression levels
400
350
300
250
200
150
100
50
0
E9.5
E11.5
E13.5
Dev. stage


H
Dnmt3a
Dnmt3b
Expression levels
800
700
600
500
400
300
200
100
0
E9.5
E11.5
E13.5
Dev. stage


I
Tubb3
Sox2
Expression levels
2500
2000
1500
1000
500
0
E9.5
E11.5
E13.5
Dev. stage

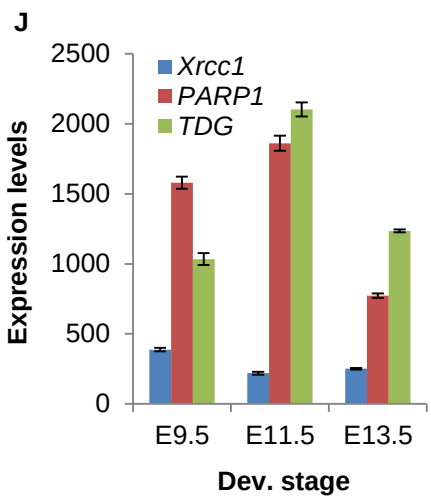

J
Xrcc1
PARP1
TDG
Expression levels
2500
2000
1500
1000
500
0
E9.5
E11.5
E13.5
Dev. stage

Figure S4 related to Figure 2. Cell populations of embryonic brain start to differ in their 5hmC content between 12.5 and 13.5 dpc. (A) 5hmC and 5mC immunostaining of sagittal sections of adult murine brain. Two areas containing grey matter regions and white matter tracts are shown. Merge views are presented. RGC - retrosplenial granular cortex, cc - corpus calossum, or - oriental layer, Py - pyramidal layer of hippocampus. Position of a region analogous to the ones shown in B and C-adult brain is marked with a dotted rectangle. (B) Co-detection of 5hmC with NeuN and DAPI in the area of adult mouse brain located at the border of grey matter region and white matter tract. NeuN-positive post-mitotic neurons exhibit high levels of 5hmC staining. Single channels and merge views are shown. (C-F) 5mC and 5hmC immunostaining of representative regions of embryonic forebrain at 12.5 dpc (left panel in C, D) and 13.5 dpc (central panel in C, E) compared with immunostaining of adult mouse brain area located at the border of grey matter region and white matter tract (right panel in C, F). Insets (D-F) and corresponding 2.5XD signal intensity plots (right panels) are presented. Scale bars are 15 μm. (G-J) Normalized mRNA expression levels of *Tet1/2/3* (G), de novo methyltransferases (H), neuronal (*β-III tubulin, Tubb3*) and NSCs' (*Sox2*) markers (I), and components of BER pathway (J) in mouse embryonic brain at the indicated stages of development extracted from GEOdatasets record GDS3442. Mean values ± standard deviations are shown.

Supplemental Experimental Procedures

Cell lines and ESCs culture. Mouse HM1 ESCs were maintained on gelatine-coated dishes in either serum-containing media (Glasgow's Minimum Essential medium (GIBCO), supplemented with 15% fetal bovine serum, 55 μM -mercaptoethanol (β- ME - GIBCO), 2 mM L-glutamine, 0.1 mM MEM non-essential amino acids (NEAA), 5 000 units ml^{-1} penicillin/streptomycin and 1 000 units ml^{-1} of murine LIF (Chemicon)) or in 2i media (Ying et al., 2008) (serum-free N2B27 (Cat. DMEM/F12: GIBCO 21331; Neurobasal: GIBCO 21103; N2: Stem Cells SF-NS-01-005; B27: GIBCO 17504-044) supplemented with 10^3 U/ml LIF, Mek inhibitor PD0325901 (1 μM) and Gsk3β inhibitor CHIR99021 (3 μM)) under feeder-free conditions. For synchronization of the cells at G1/S boundary, they were treated with 2 μg/ml of aphidicolin (Sigma Chemical Co, St Louis, MO) for 24 h. HUES-7 hESCs were maintained on Matrigel in feeder-free conditions in conditioned medium (CM) and trypsin passaged. To prepare CM, mouse embryonic fibroblasts (MEFs) were mitotically inactivated with mitomycin C (MMC) (10 μg/ml, 2.5 h) and seeded at 4.8×10^6 cells per T75 flask. After 24 h, inactivated MEFs were incubated with DMEM-F12 supplemented with 15% KnockOut Serum Replacement, 100 μM β-ME, 2 mM GlutaMAX, 1% NEAA, and 4 ng/ml βFGF for 24 h. CM was then harvested and supplemented with an additional 4 ng/ml βFGF. Human fibroblast cell line SBTeloMyoD (Hamshere et al. 1997) derived from a skin sample of a normal individual and transformed with lentivirus was maintained on DMEM (GIBCO) with 10% bovine serum.

NSCs derivation, culture and differentiation. Derivation of NSCs from lateral ventricles was performed as described (Alcock and Sottile, 2009). For fetal NSCs derivation fetal brains were isolated and washed in PBS before maceration and mechanical disruption by repeated pipetting using a glass Pasteur pipette. Cells were briefly centrifuged and washed in PBS

before plating in NSC medium containing DMEM/F12 and Neurobasal medium (1:1), N2, B27, Pen/Strep, bFGF (20 ng/ml) and EGF (20 ng/ml, Sigma). NSCs from adult or fetal origin were grown in NSC medium, using Accutase for passaging. For differentiation, NSCs were seeded in parallel on Matrigel (BD), and maintained in NSC medium (undifferentiated control), pro-glial medium (DMEM/F12 and Neurobasal medium (1:1), N2, B27, Pen/Strep, 5% FCS) or pro-neurogenic medium (DMEM/F12 and Neurobasal medium (1:1), N2, B27, Pen/Strep, 0.5% FCS, 1μM retinoic acid (Sigma), 10 g/ml BDNF). Cell culture reagents were from Life Technologies. Medium was changed every other day before fixation.

Mass spectrometry. Genomic DNA samples were digested to nucleosides by treatment with a mixture of nuclease P1 (Sigma-Aldrich), shrimp alkaline phosphatase (NEB), and DNase I (NEB), overnight at 37 °C, following a previously reported protocol (Hashimoto et al., 2014). LC-MS/MS analysis was performed in duplicate by injecting digested DNA on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A Triple Quadrupole Mass Detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out using a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of methanol and 10 mM aqueous ammonium formate (pH 4.4). MS data acquisition was performed in the dynamic multiple reaction monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram associated with its specific MS/MS transition: dC [M+H]+ at m/z 228 →112, dmC [M+H]+ at m/z 242→126, dhmC [M+H]+ at m/z 258→142, dfC [M+H]+ at m/z 256→140, and dcaC [M+Na]+ at m/z 294→178. External calibration curves with known amounts of the nucleosides were used to calculate their ratios within the samples analyzed. LC-MS/MS quantification was validated using E14 mouse embryonic stem cell DNA (Ito et al., 2011).

Immunocytochemistry, immunohistochemistry and dot blots were performed as described (Ruzov et al., 2011). Formaldehyde-fixed paraffin-embedded sections of CD1 murine

embryos and adult brain were used for immunohistochemistry. For NeuN co-localization experiments formaldehyde-fixed microtome sections were processed in whole-mount staining procedure. The samples were incubated in 2N HCl for 1 h at 37 °C.

Antibodies. Anti-5mC mouse monoclonal (Diagenode, clone 33D3,1:200 dilution), anti-5hmC mouse monoclonal (Active Motif, 1:5000 dilution), anti-5hmC rabbit polyclonal (Active Motif, 1:5000 dilution), anti-5fC rabbit polyclonal (Active Motif, 1:500 - 1:1000 dilutions), anti-5caC rabbit polyclonal (Active Motif,1:1000 - 1:3000 dilutions), anti-NeuN mouse monoclonal (Chemicon, 1:400 dilution), anti-nestin mouse monoclonal (DSHB, 1:200 dilution), anti-β-III tubulin mouse monoclonal (Berkeley Antibody Company, 1:200 dilution), chicken polyclonal anti-GFAP (Thermo Scientific, Pierce, 1:400 dilution), goat polyclonal anti-Nanog (R&D Systems, 1:50 dilution) and goat polyclonal anti-Oct4 (Santa Cruz, 1:200 dilution) primary antibodies were used. Peroxidase-conjugated anti-rabbit secondary antibody (Dako) and the tyramide signal enhancement system (Perkin Elmer, 1:200 dilution, 2 min incubation with amplification reagent) were employed for 5fC, 5caC and 5hmC detection. In 5caC co-localization experiments 5mC and 5hmC were visualized using Alexa 555-conjugated anti-mouse secondary antibody (Life Technologies).

Confocal microscopy and image analysis. Images (400-600 nm optical sections) were acquired with a Zeiss LSM 700 AxioObserver confocal microscope using a Plan-Apochromat 63x/1.40 Oil DIC M27 objective and processed using Image J and Adobe Photoshop. 3D images, signal intensity plots and profiles were generated using ZEN Zeiss LSM 700 imaging software. Co-localization coefficients were determined using the inbuilt analysis function of ZEN with a threshold intensity of 50. Signal intensity quantification was performed using Fiji software. Mean intensities were measured for at least 20 random cell nuclei for each condition. Individual data were plotted using Prism displaying mean ± s.e.m. The

significance was determined by one way ANOVA and post hoc Dunnett test, ***p<0.0001, **p<0.001, *p<0.05. Confocal raw data are available upon request.

Dot blots and immunochemistry competition experiments were performed as described (Ruzov et al., 2011) using 5hmC- and 5mC-enriched PCR fragments, 5caC DNA standard (Active Motif) and 5fC-containing oligonucleotides (gift from Active Motif). 5hmC rabbit polyclonal (Active Motif, 1:50 000 dilution), 5fC-rabbit polyclonal (Active Motif, 1:1000 – 1:3000 dilutions) and 5caC rabbit polyclonal (Active Motif, 1:3000 - 1:10 000 dilutions) antibodies were used for dot blots. Equal dilutions of DNA were loaded onto membranes. The dilution rate between two neighbouring experimental points equalled 3x in the dot blot with DNA isolated from hESCs and adult mouse brain.

Analysis of gene expression in embryonic brain. The normalized mRNA expression levels were downloaded for the entire GEOdatasets record GDS3442. Profiles for *Tet1* (Affymetrix id 1429448_s_at), *Tet2* (Affymetrix ids 1438781_at and 1455300_at), *Tet3* (Affymetrix id 1439324_at), *Xrcc1* (Affymetrix id 1416587_a_at), *PARP1* (Affymetrix ids 1422502_at and 1422503_s_at), *TDG* (Affymetrix id 1435715_x_at), *Dnmt3a* (Affymetrix ids 1460324_at and1423063_at), *Dnmt3b* (Affymetrix id 1449052_a_at), *β-III-tubulin* (Affymetrix id 1415978_at) and *Sox2* (Affymetrix id 1416967_at) were extracted for the 6 E9.5, 4 E11.5 and 6 E13.5 samples of mouse embryonic brain. Mean expression levels and standard deviations were calculated with Microsoft Excel.

Bisulfite sequencing. Genomic DNA was isolated according to standard procedures and bisulfite-treated using EpiTect bisulfite kit (Qiagen). PCR products were amplified from converted DNA and cloned to pGemT-Easy vector. 10 random clones were sequenced for each PCR product. The following primers were used:

Galanin: GGTAGTTTTTATTGGGTATAAATAG and

ATAAAACACAAATACCTTAAACAAC;

Atp5h: TTAGTTTGGAATGGGTTTTAATATT and AACCACCAAACAAACCCTAAAC;

GFAP: TTAGTAGAGGTAGGGTAGGATGGAG and

TATCTCCTTAAAACCAACATTAAAC;

Fgfr4: TTAGATTAGAAGTGGTTTAGGAGGG and

AACTAACCCAAAACTAAAAAACTCC;

5mC-, 5hmC- and 5caC-DNA IP (DIP). Genomic DNA was isolated according to standard procedures and sonicated using Diagenode Bioruptor Standard UCD-200. Sonicated DNA was purified using Qiagen DNA purification kit. 2 μg of genomic DNA was used for immunoprecipitation. 5mC-, 5hmC- and 5caC-DIP was carried out as described (Szulwach et al., 2011b; Shen et al., 2013) using mouse monoclonal 5mC (clone 33D3, Diagenode), rabbit polyclonal 5hmC (Active Motif) or rabbit polyclonal 5caC (Active Motif) antibodies and magnetic anti-rabbit Dynabeads (Invitrogen). Samples were purified using Qiagen DNA purification kit and analysed by real-time PCR. The specificity of the antibodies was assessed in DIP experiments where 1 pg of 5hmC-, 5mC-enriched, 5caC-containing or unmodified PCR fragments with identical sequence was spiked with 2 μg of genomic hESCs DNA. 5caC containing oligonucleotides were synthesised by Gene Link (Hawthorne, NY, USA) and carried 3 5caC residues (marked as X in the sequence):

CGGACCAGGCTCAGAGGTATTGGGGATCTCCCCATGTCXGCCXGCATAXGAGTT

CTGCGGAGGGATGGCATACTGTGGACCTCAGGTTGGACTGGGCCTAG. The following oligonucleotides were used for the amplification of synthetic enriched fragments: forward (GACCAGGCTCAGAGGTATTGG) and reverse (CTAGGCCCAGTCCAACCTG). DIP results were analysed by real-time PCR (RT-PCR) carried out with SYBR Green PCR Master Mix (Sigma) according to standard procedures. Fold enrichment was calculated as 2^{ddCt}, where dCt=Ct(enriched)-Ct(input) and ddCt=dCt – Ct (no antibody). Experimental error is expressed as s.e.m. The following primers were used for RT-PCR

Neuroglobin: GCCTCGGGTAGTACCTTCCT and TTTACTGGTCCAGCCCAGAG;

Galanin: GACTGTGGGTGATCCTCTCC and GGCTGGATGGTCGCTTACT;

Atp5h: GGGGAAGAGAGTGGACTCAA and CTTCGTGACCACCAGACAGA;

Pygm: CTTAGCCGGAGTGGAAAATG and GTAGCACCTTGGGGTCCTTT:

GFAP: GGACTTGGTTAGAGGCTTACA and GCTTTCCACTCTCTGCTTTC;

Olig1: ACATTTCCAGACTTCTCTC and TTTAATTGCCAGGGAGTG;

Olig2: AACTGCCACTAAGTAGAA and ATCAAACTGTGACCATTT;

Fgfr1: CCAACAAACCTTTCCTGAA and GCCTAGCAGTTGTCTCTT;

Fgfr3: CAGGTTAGTGTGAAGGAC and ACTTCCCGCATATCTCAC;

Msi1: TTCTTACCAGTTGGAAGG and TCCTCGCTATTGAGAATT;

Fgfr4: TCTCTGATGTTCTGTGAT and CTACCTTGCTCTATCCTT;

Pdgfr: AATCCTTCTGCCATCAAG and TCGCTTCTTCTATCTTCATAT

Xlr: CTCCTTTTTGCTGCTTCGAT and GAACTTCTCGGCTTCCTCAA.

TDG expression analysis. TDG expression was analysed by real-time PCR using TaqMan Gene Expression Assay (Life Technologies) for TDG (Mm02602088_g1) and was normalized for GAPDH (Mm03302249_g1). Mouse monoclonal anti-TDG (Santa Cruz) antibody was used for Western blot. TDG expression was normalized for α-tubulin with mouse monoclonal antibody (Abcam).

RT-PCR gene expression analysis was carried out with SYBR Green PCR Master Mix (Sigma) according to standard procedures. Gene expression was normalized for *Hypoxanthine-guanine phosphoribosyltransferase* (*HPRT*). The following primers were used for RT-PCR

GFAP: TGTGGATTTGGAGAGAAA and CGAACTTCCTCCTCATAG;

Olig1: TGTATGAGCTGGTGGGTTAC and GAAGGGATGCGGTGGAAG;

Olig2: TGGCTTCAAGTCATCTTC and CTGCTTCTTGTCTTTCTTG;

Glast: CCTGAACTTTGGACAGATTA and CTGCAATGATGAGTGTGA;

Fgfr1: GTCCGTCTCTTTCTCCTC and AGCGATACCACCTACTTC;

Fgfr2: CCTGGTGGAGAATGAATAC and CAGACAAACTCCACATCC;

Msi1: CTCCAAAACAATTGACCC and GCTCGAAATAGTGTTTCAC;

Fgfr3: CTACGTCACTGTACTCAA and CCTCATCAGTTTCCATCA;

Fgfr4: CGAGATGGAGGTGATGAA and GGCACATTCCACAATCAC;

Pdgfr: ACCGATTCCTTAGTCATC and CAAGGTTACTTGAGTCTC;

HPRT: GAACCAGGTTATGACCTA and TCTCCTTCATGACATCTC.

Supplemental References

Hamshere, M.G., Newman, E.E., Alwazzan, M., Athwal, B.S., Brook, J.D. (1997). Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. Proc. Natl. Acad. Sci. U S A 94, 7394-7399.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.